صادقی گورجی، شهربانو؛ پوراحمد، علی‌اکبر؛ حاجی زین‌العابدینی، محسن؛ ضیایی، ثریا(۱۳۹۴). ارزیابی کارآمدی گوگل پژوهشگر در بازیابی اطلاعات نویسندگان دارای شکل‌های گوناگون نام: بررسی ضریب بازیافت و دقت. *پژوهش‌نامه کتابداری و اطلاع‌رسانی*، ۵ (۱)، ۲۱۶-۲۰۵.

ارزیابی کارآمدی گوگل پژوهشگر در بازیابی اطلاعات نویسندگان دارای شکل‌های گوناگون نام: بررسی ضریب بازیافت و دقت

شهربانو صادقی گورجی[۱]، علی‌اکبر پوراحمد[۲]، محسن حاجی زین‌العابدینی[۳]، ثریا ضیایی[۴]

تاریخ دریافت: ۱۳۹۲/۵/۱۸ تاریخ پذیرش: ۱۳۹۲/۱۲/۴

چکیده

هدف: *هدف این مقاله بررسی قابلیت و توانایی موتور جستجوی گوگل پژوهشگر در بازیابی اطلاعات و انتشارات نویسندگان پر استناد براساس هر نوع نگارش نام است.*

روش: *این پژوهش از نوع کاربردی بود و به روش پیمایشی-توصیفی انجام شد. روش گردآوری اطلاعات از نوع میدانی و کتابخانه‌ای بود. جامعه نمونه شامل ۱۰ نویسنده پر استناد غیر ایرانی از ۵ رشته علمی بود که از طریق نمونه‌گیری غیر-احتمالی از سایت "این-سایت" انتخاب شدند. نتایج بازیابی انواع نگارش‌های نام هر نویسنده در گوگل پژوهشگر با سیاهه وارسی فهرست انتشارات وی مطابقت داده شد. برای هر نویسنده بر مبنای نگارش‌های مختلف نام به همراه استفاده از عملگر OR و محدودسازی با موضوع، ۱۲ متغیر جستجو در نظر گرفته شد که در این پژوهش از آن‌ها به‌عنوان متغیر یاد می‌شود.*

یافته‌ها: *نتایج میانگین ضریب بازیافت ۱۲ متغیر (متغیرهای به کاررفته در جستجو) برای ۱۰ نویسنده نشان داد که تنها در ۴ مورد ضریب بازیافت بالاتر از سطح متوسط و در تمام متغیرها ضریب دقت پایین‌تر از سطح متوسط است. اولویت‌بندی میانگین ضریب بازیافت و دقت متغیرها با آزمون فریدمن نشان داد که هیچ متغیری در خصوص بازیافت و دقت بر دیگر متغیرها برتری ندارد. نتایج آزمون t استودنت مستقل برای سنجش ارتباط میان متغیرهای ۱-۶ (نگارش‌های مختلف نام بدون محدودسازی با موضوع) و ۷-۱۲ (نگارش‌های مختلف نام با محدودسازی موضوعی) نشان داد که در مورد بازیافت میان این دو دسته متغیر تفاوتی وجود ندارد اما ضریب دقت در متغیرهای ۱-۶ (بدون محدودسازی نام با موضوع) بالاتر از ترکیب و محدودسازی نام با موضوع است. به‌طورکلی نتایج حاصل، حاکی از آن بود که گوگل پژوهشگر در بازیابی نگارش‌های مختلف نام نویسندگان معمولاً به‌طور مطلوب عمل نمی‌کند.*

کلیدواژه‌ها: *گوگل پژوهشگر، نویسندگان پر استناد، بازیابی اطلاعات، نام‌های نویسندگان، نگارش‌های مختلف نام نویسندگان.*

---

۱. دانشجوی کارشناسی ارشد دانشگاه پیام نور مشهد amsmath@gmail.com

۲. عضو هیات علمی گروه علم اطلاعات و دانش شناسی، دانشگاه آزاد اسلامی واحد شیروان pourahmadali@iau-shirvan.ac.ir

۳. عضو هیات علمی گروه علم اطلاعات و دانش شناسی دانشگاه شهید بهشتی zabedini@yahoo.com

۴. عضو هیات علمی دانشگاه پیام نور مشهد soraya.ziaei@gmail.com

مقدمه

نویسندگان و پژوهشگران، عوامل اصلی تولید اطلاعات در جوامع علمی به شمار می‌آیند. علی‌رغم اهمیت بسیار زیاد دسترسی به اطلاعات صحیح، جامع، روزآمد و مرتبط، دستیابی به اطلاعات مربوط به یک پژوهشگر، بسیار مشکل بوده و به‌راحتی نمی‌توان با استفاده از اسمی که در اختیار است، اطلاعات مربوط به فرد، از جمله تولیدات علمی وی را بازیابی کرد. نشر الکترونیکی و روزافزون مقالات و ورود اطلاعات آن‌ها به پایگاه‌های فراداده بدون بررسی و نظارت لازم بر اطلاعات کتابشناختی، به دشواری‌های بازیابی اطلاعات بیش‌ازپیش دامن می‌زند. ناشران بر صحت ورود اطلاعات منابع، حساسیت و نظارت چندانی ندارند و نام‌ها غالباً یا با نام کامل و یا حروف آغازین نوشته می‌شوند. همچنین تنوع نگارش نام، به‌وسیله نویسندگان به دلیل ویژگی‌های زبانی و فرهنگی و اشتباهات تایپی، جزء امور متداول است. افزایش دسترسی به منابع اطلاعاتی ناهمگون، و تنوع نگارش نام‌ها به دلیل ویژگی‌های زبانی و فرهنگی از سویی، و ناتوانایی پایگاه‌ها در تشخیص ارتباط و ایجاد پیوند میان آن‌ها - به‌نحوی‌که تمامی آثار یک نویسنده بازیابی شود یا نویسندگان همنام را به‌نوعی از هم تفکیک کند –از سویی دیگر گسیختگی و نبود انسجام را موجب شده است. فتاحی(Fattahi, 2007) این مسئله را با این پرسش مطرح می‌کند که «[آیا] مهار مستند نام‌ها در سطح جهانی عملی است؟«

با توجه به اهمیت صحت بازیابی اطلاعات مربوط به پژوهشگران و نویسندگان مقالات و دیگر منابع اطلاعات علمی، انتظار می‌رود محیط شبکه‌ای و وب، از ابزارهای ساختارمندِ قوی برای ذخیره‌سازی، دسته‌بندی و در نتیجه، بازیابی اسامی این افراد برخوردار بوده و ضریب بازیافت و دقت لازم برای بازیابی اطلاعات با جستجوی نام اشخاص را دارا باشد. با توجه به تلاش‌های زیادی که در زمینه دسته‌بندی و سازماندهی اطلاعات علمی صورت گرفته است، مطالعات کمی در مورد نویسندگان مقالات و پژوهشگران در سطح دنیا انجام شده است و حتی در داخل پایگاه‌ها نیز، تنوع در ضبط نام یک نویسنده به چشم می‌خورد. این امر ابهامات و خطاهایی را هم در جهت شناسایی نویسندگان و هم برای بازیابی منابع اطلاعاتی مرتبط با آن‌ها در وب ایجاد کرده است. کاربران بسیاری، به دلایل مختلف مثل دسترسی آسان‌تر، شناخت بیشتر، و رایگان بودن، به جستجو از طریق موتورهای کاوش عمومی، رغبت بیشتری دارند. در این میان گوگل پژوهشگر به‌عنوان زیرمجموعه گسترده‌ای از گوگل و دروازه دسترسی به اطلاعات علمی، چیزی شبیه به نمایه استنادی علوم را به ذهن متبادر می‌سازد. از آنجایی که جستجوهای استنادی به شناسایی نویسندگان و آثار کلیدی، همچنین بازیابی منابع جدید کمک می‌کنند، توجه به عناصر سازماندهی اطلاعات از جمله مهار مستند نام‌ها در این محیط‌ها اهمیت زیادی دارد. درحالی‌که

استفاده نکردن از ابزارهای مناسب، در مهار مستند نام‌ها، این موتور جستجوی محبوب را با چالش‌هایی مواجه ساخته است. بنابراین انتظار می‌رود ابزارهای جستجو در سطح وب برای بهبود وضعیت در جهت هوشمندسازی هرچه بیشتر جستجو گام بردارند.

هدف کلی این مقاله بررسی میزان کارآمدی موتور جستجوی گوگل پژوهشگر در بازیابی اطلاعات مربوط به نویسندگان است. با توجه به این هدف کلی، پژوهش حاضر چند هدف فرعی را نیز دنبال می‌کند:

۱.شناسایی توانایی گوگل پژوهشگر در برقراری ارتباط بین نگارش‌های مختلف نام نویسندگان و گردهم‌آوری آثار یک نویسنده تحت هر نگارشی از نام.

۲.شناسایی تدابیری که گوگل پژوهشگر برای تشخیص افراد دارای نام یکسان به‌کار برده است.

با توجه به متداول بودن استفاده از این موتور و لزوم شناسایی پژوهشگران هسته¬ هر علم که نقش کلیدی در نشر آثار مرتبط در همان رشته علمی ایفا می‌کنند، توانایی برقراری ارتباط بین نگارش‌های مختلف یک نام و تمییز نام‌های مشابه در این موتور جستجو که یکی از رسالت‌هایش تحلیل استنادی است، اهمیت به سزایی می‌یابد. ضرورت این امر زمانی روشن می‌شود که کاربر برای یافتن منابع منتشرشده توسط یک پژوهشگر هسته به این پایگاه مراجعه می‌کند و انتظار دارد که این موتور جستجو بتواند تسهیلاتی را فراهم کند که دقیق‌ترین یافته‌ها را در سریع‌ترین زمان ممکن به دست آورد. لذا پس از تایپ نوع خاصی از نگارش نام با مجموعه‌ای از یافته‌های درهم و بدون دسته‌بندی مواجه می‌شود که در مواردی چند مرتبط با موضوع مورد مطالعه او نیست. این مسئله نارضایتی از نتایج حاصل از جستجو را در بردارد، درحالی‌که در یک نظام بازیابی کارآمد، کاربر از میزان بازیافت و دقت نتایج رضایت دارد.

پیشینه پژوهش

همزمان با ایجاد گوگل پژوهشگر و تقریباً از اواخر سال ۲۰۰۴ افراد زیادی بر¬مبنای مطالعات تجربی به بررسی نقاط قوت و ضعف و شکاف‌های پوششی موجود در این موتور جستجو پرداخته‌اند. تمام مطالعات نیز براساس نمونه‌های بزرگ نبوده، حتی برخی از مطالعات بر¬اساس یک یا چند کلیدواژه محدود انجام شده است. از نظر نوروزی (Noruzi, 2005) ویژگی مهم گوگل پژوهشگر این است که پژوهشگر می‌تواند از آن برای ردیابی ارتباط درونی نویسندگانی که به مقالات یک موضوع خاص استناد می‌دهند، همچنین تخمین فراوانی استناد به یک مقاله خاص با عبارت «استناد شده به وسیله» استفاده کند.

برخی از پژوهشگران مزایا و معایب گوگل پژوهشگر را برشمرده و پیشنهادهایی را نیز ارائه

می‌دهند. پیتر جکسو (Jacso, 2005, 2008, 2009, 2010) از جمله نویسندگانی است که در مقالات مختلف گوگل پژوهشگر را نقد می‌کند. برخی از مهم‌ترین انتقادات وی در آثار مختلف را می‌توان به شرح زیر خلاصه کرد:

- هنوز راهکاری برای برقراری ارتباط بین نگارش‌های مختلف نام یک نویسنده و تفکیک نویسندگان نام‌های مشابه وجود ندارد.
- گرچه نام از سازمان و زمینه موضوعی تفکیک‌شده اما هنوز مشاهده می‌شود که با جستجوی نام، کلیدواژه فیلدهای دیگر هم جستجو شده است.
- هیچ راهنما یا اشاره خیلی جزئی برای اینکه چه نوع واژه خاصی باید برای زمینه مورد علاقه وارد کرد وجود ندارد. به همین علت، ناهمگونی در ورود اطلاعات گریزناپذیر است.
- در نتایج بازیابی نویسندگانی که نمایه نویسندگان خود را تکمیل کرده‌اند هنوز هم نحوه ارائه بازیابی در حالت سنتی گوگل پژوهشگر به چشم می‌خورد، به‌نحوی‌که به‌جز لینک به اطلاعات نمایه نویسندگان هیچ نشانه دیگری از استخراج اطلاعات از این نمایه در آن به چشم نمی‌خورد.

برخی دیگر نیز به کاربرد گوگل پژوهشگر در کتابخانه‌ها پرداخته‌اند. برای مثال، کسلمن و واتسن (Kesselman and Watsen, 2005) به بررسی تأثیر گوگل پژوهشگر بر سواد اطلاعاتی پرداختند. دونلن و کوک (Donlan and Cooke, 2006) به بررسی تأثیر گوگل پژوهشگر بر تغییر رفتار اطلاع‌یابی کاربران در استفاده از فهرست‌های کتابخانه‌ای پرداختند. پاتر (Potter, 2008) نیز در پژوهش خود به بررسی رابطه میان گوگل پژوهشگر و کتابخانه دانشگاهی پرداخت.

احتمالاً جامع‌ترین مطالعه تجربی موجود، مربوط به نیوهاوس (Neuhaus, 2006) است که در آوریل و جولای ۲۰۰۵ و تنها چند ماه پس از ایجاد گوگل پژوهشگر انجام شد. او ۴۷ پایگاه در تمامی موضوعات علمی را بررسی کرد. در این مطالعه ۲۳۵۰ مقاله براساس سال نشر و زبان انتخاب شدند. نتایج نشان داد که گوگل پژوهشگر در پوشش پایگاه‌های پزشکی و علوم، پایگاه‌های دسترسی رایگان و تک ناشران قدرتمند عمل می‌کند اما، برای پایگاه‌های علوم انسانی و اجتماعی دارای ضعف بوده و سوگیری به زبان انگلیسی نیز وجود دارد.

در منابع زیادی نیز این موتور جستجو با پایگاه‌های دیگر مقایسه شده است. برای مثال، مهو و یانگ (Meho and Yang, 2007) که گوگل پژوهشگر را در مقایسه با وب آو ساینس و اسکاپوس مورد بررسی قرار دادند. والترز (Walters, 2007) نیز محتوای گوگل پژوهشگر را با محتوای هفت پایگاه

دیگر مقایسه کرد. همچنین مطالعه لوین کلارک و کراوس (Levine-Clark and Kraus, 2007) گرچه رویکردی مقایسه‌ای داشت، اما یکی از پنج جستجوی خود را به جستجوی شکل‌های مختلف نام یک نویسنده اختصاص داد.

بسیاری از مقالات، موتور جستجوی گوگل پژوهشگر را براساس تحلیل استنادی مورد بررسی قرار داده‌اند، مثلاً هارزینگ و وندروال (Harzing and van der Wal, 2008) که ضریب تأثیر مجلات را بر-اساس معیارهای جی-ایندکس، اچ-ایندکس و تعداد استناد به هر مقاله مطالعه کرده‌اند.

در بین منابع موجود تنها برخی از آن‌ها بخش کوچکی از کار خود را به مسئله جستجوی نویسندگان اختصاص داده‌اند که عمدتاً به روش‌های جستجوی نام نویسندگان و یا به‌طورکلی کاستی‌های این موتور جستجو در بازیابی نام نویسندگان اختصاص دارند. در این میان می‌توان به مقاله لوین کلارک و کراوس (Levine-Clark and Kraus, 2007) و جکسو (Jacso, 2009) اشاره کرد که به بررسی بازیابی‌های نه چندان مرتبط گوگل پژوهشگر براساس فیلد جستجوی نام نویسنده می‌پردازد.

رویکردهای مطالعاتی موجود در منابع خارج از کشور در منابع داخلی نیز قابل مشاهده است، لذا افزون بر منابع چندی که به بررسی ویژگی‌های گوگل پژوهشگر پرداخته‌اند، منابعی نیز وجود دارند که رویکرد مطالعه مقایسه‌ای و استنادی را دنبال می‌کنند. به عنوان نمونه، رمضانی (Ramezani, 2007) براساس مقایسه‌های مختلف گوگل پژوهشگر با دیگر پایگاه‌های اطلاعات علمی اذعان می‌دارد که دو مورد به گوگل پژوهشگر اعتبار ویژه‌ای داده است: اول جستجوی کلیدواژه‌ای در متن کامل مقاله‌ها و دیگر پیوند با مقاله‌هایی که بعداً به آن استناد کرده‌اند. وی نیز به نادرستی نتایج جستجوی گوگل پژوهشگر در تشخیص نام نویسندگان اشاره کرده است.

همان‌طور که بیان شد رویکرد مطالعاتی حاکم بر پیشینه‌های خارجی و داخلی درباره گوگل پژوهشگر، بیان ویژگی‌ها، کاربرد در کتابخانه‌ها، مقایسه با پایگاه‌های دیگر و تحلیل، استنادی است و در بسیاری از مقالات، پیشنهادهایی برای بهبود عملکرد گوگل پژوهشگر مطرح شده است. در این پژوهش‌ها از روش‌های مختلفی چون مشاهده، تحلیل محتوای نتایج بازیابی و روش مقایسه نتایج بین پایگاه‌های مختلف استفاده شده است. برخی از پژوهش‌ها برمبنای جامعه پژوهش محدود و برخی دیگر گسترده انجام شده است؛ ولی نکته حائز اهمیت، نتایج تقریباً مشابه آنهاست.

روش پژوهش

پژوهش حاضر از نوع کاربردی بوده و با استفاده از دو روش مطالعه کتابخانه‌ای و اسنادی و روش

پیمایشی انجام شده است. جامعه پژوهش را تمامی پژوهشگرانی تشکیل دادند که در پایگاه این¬سایت به‌عنوان پر استنادترین پژوهشگران معرفی شده‌اند. از این بین ۱۰ نویسنده پر استناد از ۵ رشته علمی به روش غیر احتمالی انتخاب شده و متغیرهای جستجوی نام آن‌ها با الهام از لوین کلارک و کراوس-Levine) (Clark and Kraus, 2007طراحی و جستجو بر طبق آن‌ها انجام شد که برای اسامی سه وجهی ۱۲ متغیر جستجو و اسامی ۲ وجهی ۸ متغیر بررسی شد. در پژوهش حاضر از هر متغیر به‌عنوان یک متغیر یاد می‌شود. مرحله بعد، مطابقت نتیجه جستجوی هر یک از حالت‌ها با فهرست منابع هر شخص و تعیین ضریب بازیافت و دقت آن‌ها براساس متغیر زیر بود.

$$\text{نسبت بازیافت} = \frac{\text{تعداد پیشینه‌های مرتبط بازیابی شده}}{\text{تعداد پیشینه‌های مرتبط موجود}} * ۱۰۰$$

$$\text{نسبت دقت} = \frac{\text{تعداد پیشینه‌های مرتبط بازیابی شده}}{\text{تعداد پیشینه‌های بازیابی شده در جستجو}} * ۱۰۰$$

از آنجایی که تعیین تعداد پیشینه‌های مرتبط موجود در پایگاه غیرممکن بود، پژوهشگر با فرض اینکه هر عنوان حداقل یک‌بار در پایگاه نمایه شده است (به‌عنوان استناد یا اصل منبع)، ابتدا تمامی منابع تا سال ۲۰۱۰ را در گوگل پژوهشگر جستجو نمود تا کل منابع به‌عنوان تعداد پیشینه‌های مرتبط موجود در پایگاه در نظر گرفته شود و موارد تکراری به‌این‌علت که ضریب دقت و بازیافت را به‌طور کاذب تغییر می‌دادند تنها یک‌بار به‌عنوان بازیابی شده‌ی مرتبط در نظر گرفته شدند.

در این پژوهش به پرسش‌های زیر پاسخ داده شده است:

۱.گوگل پژوهشگر چه میزان مهار مستندات نام‌های اشخاص حقیقی را رعایت کرده است؟

۲.گوگل پژوهشگر چه میزان تشخیص و تفکیک آثار نویسندگان متفاوت دارای نام مشترک را رعایت کرده است؟

این پرسش در قالب ۴ پرسش فرعی زیر پاسخ داده شد:

۱.کدام متغیرها در بازیابی اطلاعات جامعه پژوهش، از میزان بازیافت بیشتری برخوردار هستند؟

۲.کدام متغیرها در بازیابی اطلاعات جامعه پژوهش، از میزان دقت بیشتری برخوردار هستند؟

۳.آیا بین متغیرهای ۱ تا ۶ (بدون محدودسازی با موضوع) با متغیرهای ۷ تا ۱۲ (ترکیب و محدودسازی با

موضوع) از نظر میزان بازیافت رابطه وجود دارد؟

۴.آیا بین متغیرهای ۱ تا ۶ (بدون محدودسازی با موضوع) با متغیرهای ۷ تا ۱۲ (ترکیب و محدودسازی با موضوع) از نظر میزان دقت رابطه وجود دارد؟

فرمول‌های استفاده‌شده که در این پژوهش به‌عنوان متغیر در نظر گرفته شده‌اند در جدول ۱ نشان داده شده است:

جدول۱. متغیرها و فرمول‌های جستجو

متغیر ۱	نام خانوادگی، نام (فرمول ۱)
متغیر ۲	نام خانوادگی، آغازه نام (فرمول ۲)
متغیر ۳	نام خانوادگی، دو وجه آغازه نام (برای اسامی سه وجهی) (فرمول ۳)
متغیر ۴	دو وجه آغازه نام + فاصله + نام خانوادگی در داخل گیومه (برای اسامی سه وجهی) (فرمول ۴)
متغیر ۵	آغازه نام + فاصله + نام خانوادگی در داخل گیومه (فرمول ۵)
متغیر ۶	تمام شکل‌های فوق که با عملگر OR ترکیب شدند (فرمول ۶)
متغیر ۷	نام خانوادگی، نام محدودشده به موضوع (فرمول ۷)
متغیر ۸	نام خانوادگی، آغازه نام محدودشده به موضوع (فرمول ۸)
متغیر ۹	نام خانوادگی، دو وجه آغازه نام (برای اسامی سه وجهی) محدودشده به موضوع (فرمول ۹)
متغیر ۱۰	دو وجه آغازه نام + فاصله + نام خانوادگی در داخل گیومه (برای اسامی سه وجهی) محدودشده به موضوع (فرمول ۱۰)
متغیر ۱۱	آغازه نام + فاصله + نام خانوادگی در داخل گیومه محدودشده به موضوع (فرمول ۱۱)
متغیر ۱۲	تمام شکل‌های فوق که با عملگر OR ترکیب شدند محدودشده به موضوع (فرمول ۱۲)

منظور از موضوع در متغیرهای مورد بررسی، رشته پژوهشی نویسندگان انتخاب شده است که براساس اطلاعات به‌دست‌آمده از پایگاه این¬سایت در فرمول‌ها به کار گرفته شد.

تجزیه و تحلیل داده‌ها

در این پژوهش به‌منظور سنجش ۹۲ متغیر (مجموع متغیرهای جستجوی نام ۱۰ نویسنده) به‌طورکلی ۵۶۲۰۴ پیشینه بازیابی شده مورد بررسی و با فهرست انتشارات نویسندگان مطابقت داده شد. جدول‌های ۲ و ۳ به ترتیب ضریب بازیافت و ضریب دقت متغیرهای مطرح برای جستجوی نام هر نویسنده را نشان می‌دهند.

جدول ۲. درصد ضریب بازیافت ۱۰ نویسنده منتخب در ۱۲ متغیر

	نویسنده ۱	نویسنده ۲	نویسنده ۳	نویسنده ۴	نویسنده ۵	نویسنده ۶	نویسنده ۷	نویسنده ۸	نویسنده ۹	نویسنده ۱۰
متغیر ۱	۴۶	۹۸	۶۷	۷۳	۸۷	۷۰	۱۶	۹۴	۷۲	۶۰
متغیر ۲	۸۴	۹۱	۷۶	۵۳	۹۵	۷۸	۱۰	۸۷	۷۵	۸۶
متغیر ۳	۸۲	----	۷۸	۶۵	۵۹	۵	۱۶	----	----	----
متغیر ۴	۸۳	----	۸۳	۶۵	۴۹	۲	۱۹	----	----	----
متغیر ۵	۸۸	۹۲	۳۹	۴۷	۹۳	۶۳	۱۹	۹۴	۷۸	۸۹
متغیر ۶	۴۴	۹۹	۴۰	۷۳	۶۷	۰	۱۸	۹۱	۷۵	۵۶
متغیر ۷	۹	۴۸	۱۶	۶۸	۶۴	۷۳	۱۵	۸۷	۶۲	۴۴
متغیر ۸	۲۹	۲۴	۴۹	۶۲	۷۸	۸۹	۱۵	۹۴	۵۸	۹۲
متغیر ۹	۲۷	----	۲۱	۶۸	۴۸	۴	۴۷	----	----	----
متغیر ۱۰	۲۸	----	۱۱	۸۰	۴۸	۲	۴۶	----	----	----
متغیر ۱۱	۲۸	۳۷	۷۵	۶۸	۷۷	۸۲	۱۵	۹۱	۸۲	۹۲
متغیر ۱۲	۹	۴۳	۲۶	۷۴	۵۳	۰	۱۶	۸۴	۶۳	۴۳

همان‌گونه که در جدول شماره ۲ مشاهده می‌شود، بیشترین میزان بازیافت مربوط به ترکیب تمام شکل‌های نام با عملگر OR (متغیر ۶) و Tibshirani, Robert (نویسنده دوم) است که با توجه به پیوست، مربوط به ترکیب تمامی نگارش‌های نام با عملگر OR بدون محدودسازی موضوعی است. درحالی‌که در متغیرهای ۷-۱۲ که هر یک از نام‌ها با موضوع محدود شدند، کاهش محسوسی مشاهده می‌شود.

کمترین میزان بازیافت نیز مربوط به ترکیب تمامی نگارش‌های نام با عملگر OR و محدودشده به موضوع (متغیر ۱۲) و Yang, Peidong (نویسنده ششم) است که ترکیب تمامی نگارش‌های نام با عملگر ORو محدودسازی با موضوع است. با دقت بیشتر در جدول و مقایسه نتایج حاصل نوعی ناهمگونی بازیابی در کل و یا حتی در نام‌های یک نویسنده خاص نمود پیدا می‌کند.

جدول ۳. درصد ضریب دقت ۱۰ نویسنده منتخب در ۱۲ متغیر

	نویسنده ۱	نویسنده ۲	نویسنده ۳	نویسنده ۴	نویسنده ۵	نویسنده ۶	نویسنده ۷	نویسنده ۸	نویسنده ۹	نویسنده ۱۰
متغیر ۱	۶۲	۱۵	۲۳	۱۰	۷۴	۳۱	۸	۵	۱۳	۷۱
متغیر ۲	۲۸	۸	۱۵	۷	۳۶	۱۶	۵	۳	۱۰	۵۲
متغیر ۳	۴۸	----	۲۰	۱۱	۲۹	۱	۸	----	----	----
متغیر ۴	۴۹	----	۲۲	۱۶	۲۴	۵	۱۰	----	----	----
متغیر ۵	۳۴	۸	۸	۶	۳۷	۱۳	۱۰	۳	۱۲	۵۸
متغیر ۶	۶۳	۱۵	۲۳	۳۱	۷۳	۰	۱۶	۵	۱۶	۶۷
متغیر ۷	۵۹	۲۴	۱۱	۲۵	۸۷	۶۸	۹	۵	۱۴	۶۳
متغیر ۸	۲۵	۶	۲۳	۹	۵۸	۱۸	۷	۳	۱۰	۷۶
متغیر ۹	۵۶	----	۱۲	۲۱	۳۹	۵	۲۷	----	----	----
متغیر ۱۰	۵۷	----	۶	۲۷	۴۰	۲۸	۲۹	----	----	----
متغیر ۱۱	۴۵	۹	۳۶	۱۱	۵۸	۱۷	۷	۳	۱۶	۷۶
متغیر۱۲	۵۹	۲۲	۱۹	۳۶	۸۶	۰	۱۸	۶	۱۸	۶۳

در جدول ۳ نیز مشاهده می‌شود که بیشترین میزان دقت مربوط به نام خانوادگی، نام محدودشده به موضوع (متغیر ۷) و Alivisatos, Paul (نویسنده پنجم) است که با توجه به پیوست، مربوط به نگارش نام خانوادگی، نام محدودشده به موضوع است. درحالی‌که همین متغیر برای دیگر نویسندگان نتیجه مشابه یا حتی نزدیک را در برنداشت. کمترین میزان بازیافت نیز مربوط به متغیرهای ۶ و ۱۲ نویسنده ششم است که ترکیب تمامی نگارش‌های نام با عملگر OR و به ترتیب بدون محدودسازی و با محدودسازی موضوعی است. لذا عملگر OR کارکرد مورد انتظار را نشان نداده است که البته با دقت در دیگر نتایج حاصل کم و بیش این مسئله تکرار شده است.

علاوه بر نتایج توصیفی ذکر شده که از مشاهده جدول‌های ۲ و۳ حاصل می‌شود برای پاسخگویی دقیق به پرسش‌های پژوهش نیاز به آمار استنباطی است.

پاسخ به پرسش‌های پژوهش

همان‌گونه که پیش‌تر بیان شد، این پژوهش به دنبال پاسخگویی به ۲ پرسش اصلی و ۴ پرسش فرعی است. به پرسش اول پژوهش، براساس میانگین نتایج برای ۱۰ نویسنده پاسخ داده شد. و از آنجایی

که پاسخ به پرسش دوم نیازمند ملزومات خاصی بود، در قالب ۴ پرسش فرعی مطرح و بررسی شد .

پرسش اول: گوگل پژوهشگر چه میزان مهار مستندات نام‌های اشخاص حقیقی را رعایت کرده است؟

با توجه به دو کارکرد عمده مهار مستند نام اشخاص و هدفی که سنجش میزان بازیافت و دقت دنبال می‌کند به پرسش اول براساس میانگین نتایج ضریب بازیافت و دقت ۱۰ نویسنده برای هر متغیر پاسخ داده شد.

جدول۴. درصد میانگین بازیافت و دقت متغیرهای جستجو برای نویسندگان

شماره متغیر	میانگین بازیافت	میانگین دقت
۱	۶۸	۳۱
۲	۷۴	۱۸
۳	۵۱	۲۰
۴	۵۰	۲۱
۵	۷۰	۱۹
۶	۵۶	۳۱
۷	۴۹	۳۷
۸	۵۹	۲۴
۹	۳۶	۲۷
۱۰	۳۶	۳۱
۱۱	۶۵	۲۸
۱۲	۴۱	۳۳

اگر ۵۰ درصد را متوسط در نظر بگیریم، همان‌گونه که در جدول ۴ مشاهده می‌شود، در ۴ متغیر، ضریب بازیافت کمتر از سطح متوسط و در بقیه موارد بالاتر از سطح متوسط قرار دارد. اما در رابطه با ضریب دقت به این علت که میانگین آن در تمام موارد پایین‌تر از سطح متوسط قرار دارد، نتیجه مطلوب نمی¬باشد. از آنجا که مهار مستندات ۲ کارکرد عمده را همزمان اعمال می‌کند، لازم است ضریب بازیافت و دقت در همه موارد و همراه با هم حداقل بالاتر از سطح متوسط قرار داشته باشند تا بتوان عملکرد لازم را تبیین شده در نظر گرفت.

پرسش دوم: گوگل پژوهشگر چه میزان تشخیص و تفکیک آثار نویسندگان متفاوت دارای نام مشترک را رعایت کرده است؟

به پرسش اصلی فوق در قالب ۴ پرسش فرعی پاسخ داده شد .

پاسخ به پرسش ۲-۱: کدام متغیرها برای بازیابی آثار نویسنده از میزان بازیافت بیشتری برخوردار هستند؟

چون بیشتر یا کمتر بودن میزان ضریب بازیافت اهمیت دارد با استفاده از آزمون ناپارامتری فریدمن متغیرهای نویسندگان در خصوص میزان بازیافت اولویت‌بندی شدند.

جدول ۵. نتایج آزمون فریدمن برای اولویت‌بندی متغیرهای ۱۰ نویسنده در خصوص میزان بازیافت

آماره	مقدار
کی دوی فرید من	۷/۱۴۷
درجه آزادی	۱۱
سطح معنی‌داری	۰/۷۸۷

همان‌گونه که در جدول ۵ ملاحظه می‌شود، سطح معنی‌داری آزمون فریدمن معادل ۰/۷۸۷ است و این مقدار از ۰/۰۵ بیشتر است، لذا فرض صفر تائید می‌شود. یعنی بین متغیرهای مجموع ۱۰ نویسنده در خصوص میزان بازیافت تفاوت معنی‌داری وجود ندارد. این بدان معناست که هیچ‌یک از متغیرها در خصوص میزان بازیافت بر دیگری برتری ندارد و نمی‌توان به کاربران یک یا چند متغیر جستجوی مشخص را پیشنهاد کرد که حداقل با آن بتوانند نتایج بیشتری را بازیابی کنند.

پاسخ به پرسش۲-۲: کدام متغیرها برای بازیابی آثار ۱۰ نویسنده از میزان دقت بیشتری برخوردار است؟

برای پاسخ به این پرسش نیز با استفاده از آزمون ناپارامتری فریدمن متغیرهای نویسندگان در خصوص میزان دقت اولویت‌بندی شدند.

جدول۶. نتایج آزمون فریدمن برای اولویت‌بندی متغیرهای ۱۰ نویسنده در خصوص میزان دقت

آماره	مقدار
کی دوی فرید من	۱۸/۸۸۲
درجه آزادی	۱۱
سطح معنی‌داری	۰/۰۶۳

همان‌گونه که در جدول ۶ ملاحظه می‌شود، احتمال سطح معنی‌داری آزمون فریدمن معادل ۰/۰۶۳ است و این میزان از ۰/۰۵ بیشتر است، لذا فرض صفر تائید می‌شود. یعنی بین متغیرهای تعریف‌شده برای ۱۰ نویسنده در خصوص میزان دقت تفاوت معنی‌داری وجود ندارد و هیچ‌یک از متغیرها از نظر بهبود میزان دقت جستجو بر دیگری برتری ندارد. لذا در این مورد نیز نمی‌توان به کاربران پیشنهاد کرد که کدام متغیر برای جستجوی نام نویسندگان دقت بهتری را خواهد داشت.

پاسخ به پرسش۲-۳ : آیا بین متغیرهای ۱ تا ۶ (بدون محدودسازی با موضوع) با متغیرهای ۷ تا ۱۲ (ترکیب و محدودسازی با موضوع) از نظر میزان بازیافت تفاوت معنی‌دار وجود دارد؟

از آنجا که برای تعیین نوع آزمون مقایسه‌ای باید از نرمال بودن متغیرها اطمینان حاصل شود سطح معنی‌داری متغیرها براساس آزمون کولوموگروف اسمیرنف بررسی شد .

جدول۷. نتایج آزمون کولوموگروف اسمیرنف متغیرهای تحت بررسی در خصوص میزان بازیافت مقیاس تحت بررسی

مقیاس تحت بررسی	میانگین	انحراف معیار	آماره Z کولوموگروف-اسمیرنف	سطح معنی‌داری	نتیجه
بدون محدودسازی با موضوع	۵۴/۲۲	۲۲/۹۶۹	۰/۷۶	۰/۶۱۱	نرمال
ترکیب و محدودسازی با موضوع	۴۲/۲۲	۱۹/۵۸	۰/۴۴۴	۰/۹۸۹	نرمال

با توجه به جدول ۷ و بالاتر بودن احتمال سطح معنی‌داری در تمام موارد از ۰/۰۵، لذا کلیه متغیرهای تحت بررسی نرمال بوده و در نتیجه از آزمون پارامتری t استودنت برای مشخص کردن تفاوت بین دو دسته از متغیرها استفاده شد.

جدول ۸. نتایج آزمون t استودنت مستقل برای متغیرهای تحت بررسی در خصوص میزان بازیافت

آماره t	درجه آزادی	سطح معنی‌داری
۱/۲۲۲	۵	۰/۲۷۶

همانگونه که در جدول ۸ مشاهده می‌شود مقدار سطح معنی‌داری برابر ۰/۲۷۶ می‌باشد که بیشتر از ۰/۰۵ است، لذا بین متغیرهای ۱ تا ۶ (بدون محدودسازی با موضوع) با متغیرهای ۷ تا ۱۲ (ترکیب و محدودسازی با موضوع) از نظر میزان بازیافت تفاوت معنی‌داری وجود ندارد. به این معنا که در صورت محدودسازی متغیرها با موضوع، نتیجه بازیافت تفاوتی نخواهد کرد. بنابراین با توجه به نتیجه حاصل در این خصوص می‌توان گفت کاربر نمی‌تواند با محدود کردن جستجوی خود از نظر موضوعی انتظار بازیابی بهتری داشته باشد .

پاسخ به پرسش۲-۴: آیا بین متغیرهای ۱ تا ۶ (بدون محدودسازی با موضوع) با متغیرهای ۷ تا ۱۲(ترکیب و محدودسازی با موضوع) از نظر میزان دقت تفاوت معنی‌دار وجود دارد؟

در این مورد هم مانند پاسخ پرسش قبل برای تعیین نوع آزمون مقایسه‌ای باید از نرمال بودن متغیرها اطمینان حاصل شود لذا سطح معنی‌داری متغیرها براساس آزمون کولوموگروف اسمیرنف بررسی شد.

جدول ۹: نتایج آزمون کولوموگروف اسمیرنف متغیرهای تحت بررسی در خصوص دقت

مقیاس تحت بررسی	میانگین	انحراف معیار	آماره Z کولوموگروف-اسمیرنف	سطح معنی‌داری	نتیجه
بدون محدودسازی با موضوع	۷۲/۶۶	۵۲/۴۱۲	۰/۷۲۳	۰/۶۷۳	نرمال
ترکیب و محدودسازی با موضوع	۳۱/۶۱	۱۹/۱۷۴	۰/۸۴۸	۰/۴۶۸	نرمال

در اینجا نیز پس از کسب اطمینان از نرمال بودن متغیرها (جدول ۹) از آزمون t استودنت مستقل (جدول ۱۰) استفاده شد.

جدول ۱۰: نتایج آزمون t استودنت مستقل برای متغیرهای تحت بررسی در خصوص دقت

آماره t	درجه آزادی	سطح معنی‌داری
۲/۰۹۱	۵	۰/۰۳۴

همانگونه که در جدول ۱۰ مشاهده می‌شود مقدار سطح معنی‌داری برابر ۰/۰۳۴ می‌باشد که کمتر از ۰/۰۵ است، لذا بین متغیرهای ۱ تا ۶ (بدون محدودسازی با موضوع) با متغیرهای ۷ تا ۱۲(ترکیب و محدودسازی با موضوع) از نظر میزان دقت تفاوت معنی‌داری وجود دارد. با توجه به جدول ۹ مشاهده می‌شود که روش بدون محدودسازی با موضوع بهتر از ترکیب و محدودسازی با موضوع در خصوص دقت است. لذا می‌توان نتیجه گرفت گوگل پژوهشگر توانایی برقراری ارتباط مناسب میان نام نویسنده و موضوع مربوط به نویسنده را ندارد. بنابراین کاربر نمی‌تواند از درستی نتایج حاصل از جستجوی خود رضایت و اطمینان داشته باشد .

به‌طور خلاصه می‌توان این‌گونه مطرح کرد که گرچه گوگل پژوهشگر سبک‌های استناد را معیاری برای جستجوی نگارش‌های مختلف نام در نظر می‌گیرد اما نمی‌تواند ارتباط میان آن‌ها را به‌خوبی تشخیص دهد و برقرار کند. چنانچه مشاهده شد اگر کاربر هر شکل از نگارش‌های نام را در جستجو استفاده کند، میزان بازیافت و دقت بهبود نخواهد یافت؛ به این معنا که با جستجوی هر شکل از نام، برخی از آثار نویسنده بازیابی و برخی دیگر بازیابی نمی‌شوند. در حالی که میزان بازیابی‌های کل افزایش می‌یابد. زیرا نویسندگان دیگر با نام‌های مشابه در نتیجه بازیابی‌ها فهرست می‌شوند. از آنجایی که تنها محدودگر بازیابی‌های نام نویسنده در گوگل پژوهشگر موضوع پژوهشی نویسندگان است مشاهده شد که ترکیب موضوع با نام‌ها نیز تأثیری بر بازیافت ندارد و هنگامی‌که موضوع با نگارش‌های مختلف نام ترکیب شود میزان دقت کاهش می‌یابد. در حالی که این نتیجه خلاف تصور منطقی است و کاربر انتظار دارد وقتی نام نویسنده را با موضوع ترکیب می‌کند دقت جستجو افزایش یابد .

بنابراین راهکارهای ارائه‌شده به وسیله گوگل پژوهشگر که همان استفاده از عملگر OR و محدودسازی موضوعی است عملی نخواهد بود و کاربر ناچار است برای بازیابی موفق‌تر تمامی نگارش‌های نام نویسنده را جستجو کند.

نتیجه

با توجه به آنچه تاکنون گفته شد، می‌توان ادعا کرد که پژوهش حاضر با بررسی ضریب بازیافت و دقت گوگل پژوهشگر در جستجوی نام ۱۰ نویسنده منتخب نتایجی را در برداشت که از سویی بر شگردهای جستجو از طرف کاربر توجه دارد و از سویی دیگر بر نحوه سازمان‌دهی اطلاعات از طرف موتور کاوش متمرکز است.

گرچه گوگل پژوهشگر سبک‌های استناد را معیاری برای جستجوی نگارش‌های مختلف نام در نظر گرفته است، اما نمی‌تواند ارتباط میان آن‌ها را تشخیص داده و یا برقرار کند. به نظر رسید اگر نام به خودی خود دارای ویژگی خاصی نبود (مثل دو وجهی یا سه وجهی بودن، ویژگی‌های زبان‌شناختی و یا متداول یا خاص بودن نام)، گوگل پژوهشگر برای برقراری ارتباط میان نگارش‌های مختلف نام یک نویسنده تدبیر خاصی نیندیشیده است. بین شکل‌های مختلف نام نویسندگان پیوندی وجود ندارد تا اگر کاربر یک شکل از نام را جستجو کند شکل‌های دیگر نام نیز در نتایج بازیابی لحاظ شود. در نتایج ۱۰ نویسنده در مجموع هماهنگی خاصی مشاهده نشد. حتی نتایج بازیابی نگارش‌های مختلف نام برخی از نویسندگان از نوسان چشمگیری برخوردار بود.

اگر کاربر به دنبال دریافت نتایج بازیابی جامع باشد و ناهمگونی نگارش نام نیز وجود داشته باشد نمی‌توان متغیر جستجوی خاصی را پیشنهاد کرد. پیشنهاد گوگل پژوهشگر در این مورد استفاده از عملگر ORدر بازیابی‌هاست؛ درحالی‌که نتایج نشان داد عملگر OR کارکرد لازم را ارائه نمی‌دهد. این نتیجه‌گیری در پژوهش جکسو (Jacso, 2009) و لوین کلارک و کراوس (Levine-Clark and Kraus, 2007)نیز حاصل شده بود.

گوگل پژوهشگر نمی‌تواند بین نام نویسندگان و موضوعات مربوط به آن‌ها ارتباط کامل برقرار سازد. لذا در مجموع نه تنها موضوع موجب بهبود نتایج بازیابی نمی‌شود بلکه در بسیاری از موارد نتایج را محدود نیز می‌سازد. گوگل پژوهشگر در رابط کاربر خود تنها از موضوعات مشخصی به‌عنوان محدودگر استفاده کرده است. بنابراین کاربر، از نظر تعداد موضوعاتی که در رابط کاربر گوگل پژوهشگر وجود دارد گزینه‌های انتخابی محدودی دارد. با توجه به جدول شماره ۳ نیز مشاهده شد که در خصوص میانگین

دقت، روش بدون محدودسازی با موضوع از ترکیبی و محدودسازی با موضوع بهتر عمل می‌کند و در مورد میزان بازیافت نیز تفاوت وجود نداشت.

زمانی که نام متداولی مورد جستجو قرار می‌گیرد، نتایج بازیابی به شکل محسوسی افزایش خواهد داشت که فقدان محدودگرهای کافی در گوگل پژوهشگر موجب سردرگمی کاربر در حجم زیاد بازیابی‌ها می‌شود. طراحان نظام‌های ذخیره و بازیابی برای ارائه نتایج بهتر راهکارهایی را ارائه می‌دهند؛ یکی از این راهکارها استفاده از محدودگرها برای کمک به کاربران است که بتوانند با استفاده از آن‌ها نتایج بهتری از جستجوی خود بیابند. بنابراین نتایج نشان می‌دهد که در گوگل پژوهشگر عناصر کافی برای تفکیک نویسنده از نویسندگان با نام مشابه لحاظ نشده است. این امر به دلیل استفاده نشدن از ابزارهای مستندسازی و الگوهای ارتباطی برای ابهام‌زدایی و ایجاد روابط میان نگارش‌های مختلف نام نویسندگان و آثار مربوط است .

از آنجایی که موتور جستجوی گوگل پژوهشگر باید از ابزارهای ساختارمندِ قوی برای ذخیره‌سازی، دسته‌بندی و در نتیجه، بازیابی اسامی نویسندگان برخوردار بوده و بازیافت و دقت لازم برای بازیابی اطلاعات اشخاص را دارا باشد، لازم است که تغییرات و گسترش‌هایی در نرم‌افزارهای گردآوری اطلاعات، رابط کاربر و نمایش نتایج بازیابی انجام دهد. گوگل پژوهشگر به‌عنوان موتور جستجویی که به‌طور خاص به گردهم‌آوری و ارائه اطلاعات و منابع دانشگاهی و پژوهشی می‌پردازد و سیاست مدیریت استنادات را دنبال می‌کند نیاز دارد تا از ابزارهایی که بتواند به‌طور سازگارتر و انعطاف‌پذیرتر عناصر اطلاعاتی را سازمان‌دهی کند، بهره گیرد و در این بین استفاده از ابزارهای مهار مستند نام اشخاص حقیقی به‌عنوان یک بازوی مهم استناد حائز اهمیت است. گوگل پژوهشگر می‌تواند با در نظر گرفتن محدودکننده‌هایی از جمله موضوع، زبان، کشور، دانشگاه و ... نویسنده، کاربر را در بازیابی هرچه بهتر یاری رساند.

References

Donlan, R. and R. Cooke, (2006). Running with the devil: Accessing library-licensed full text holdings through google scholar. Internet Reference Services Quarterly. 10(3-4): p. 149-157.

Fattahi, R (2007). From ideals to realities: a review of the most important challenges and approaches of information organization in the present time. Library and Information Science. 4 (40) (In Persian).

Harzing, A.W.K. and R. van der Wal, (2008). Google Scholar as a new source for citation analysis. Ethics in Science and Environmental Politics. 8(1): p. 61-73.

Jacso, P. (2005). Google Scholar: the pros and the cons. Online Information Review. 29(2):

p. 208-214.
Jacso, P. (2008). Savvy searching - Google Scholar revisited. Online Information Review. 32(1): p. 102-114.
Jacso, P. (2009). Google scholar's ghost authors. Library Journal. 134(18).
Jacso, P. (2010). Metadata mega mess in Google Scholar. Online Information Review. 34(1): p. 175-191.
Kesselman, M. and S.B. Watsen, (2005). Google Scholar and libraries: Point/counterpoint. Reference Services Review. 33(4): p. 380-387.
Kousha, K. (2008). Overlap between ISI and Google Scholar in Science disciplines, Quarterly Journal of National Library of Iran, 18(3), 213-234 (In Persian).
Levine-Clark, M. and J. Kraus, (2007). Finding chemistry information using Google Scholar: A comparison with Chemical Abstracts Service. Science and Technology Libraries. 27(4): p. 3-17.
Meho, L.I. and K. Yang, (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. Journal of the American Society for Information Science and Technology. 58(13): p. 2105-2125.
Neuhaus, C., et al., (2006). The depth and breadth of Google Scholar: An empirical study. Portal-Libraries and the Academy. 6(2): p. 127-141.
Noruzi, A. (2005). Google Scholar: The new generation of citation indexes. LIBRI. 55(4),p. 170-180.
Potter, C. (2008). Standing on the shoulders of libraries: A holistic and rhetorical approach to teaching google scholar. Journal of Library Administration. 47(1-2): p. 5-28.
Ramezani, A. (2007). Librarians’ questions from Google Scholar. Information Seeking and Information Sciences 5(1) 15-16 (In Persian).
Walters, W.H., (2007) Google Scholar coverage of a multidisciplinary field. Information Processing and Management. 43(4): p. 1121-1132.
Evaluation of the effectiveness of Google Scholar in authors' information retrieval
Sh. Sadeghi Gouraji , A. A. Pourahman , M. Hajizeinolabedini , S. Zeiaei